ISSN: 1992-8645

www.jatit.org

**JATIT** E-ISSN: 1817-3195

# E-COMMERCE CUSTOMER'S OPINION PREDICTION ON MULTIPOLARITY WORDS WITH RULE BASED CLASSIFICATION TECHNIQUE

A. AL FIRTHOUS<sup>1</sup>, P.ARUL<sup>2</sup>

<sup>1</sup>Research Scholar Bharathiar University Coimbatore, Tamil Nadu, India <sup>2</sup>Assistant Professor Department of Computer Science and Applications Government Arts College, Tiruchirapalli – 22 Tamil Nadu, India

E-mail: 1 alfirquest@gmail.com

#### ABSTRACT

The foremost goal of this research is to analyze the reviews of E-Commerce products using Social Media Big Data using proposed Rule Based Classification Technique. This proposed work is about evaluating customer's Tweets about Amazon products. Customer's Tweets may vary according to the person's taste and style. And Amazon reviews presented in the website is very huge this is making it impossible for customers to discover about product reaction. This issue is addressed in the proposed approach, which involves undertaking opinion mining across numerous social networks. Its goal is to gather reviews from Twitter in order to scrutinize Amazon product sentiment. And it delivers useful information to online shopping customer's favorite for and buys products. This work take into account of various kinds of opinion words found in tweets, such as capitalized words, repeated letter sequences, negation words, modifier items, emoji, intensifiers, slang terms, conjunction words and exclamatory words. This proposed work integrated the five kinds of sentiment or lexicon dictionary. This kind of amalgamation can reduce the lacking of opinion words, and also can avoid the missing polarity value.

Keywords: Opinion Mining, Analysis, Prediction, Classification, Amazon, Products

#### 1. INTRODUCTION

In the E-Commerce industry, opinion mining is critical for making timely decisions about customer retention and earnings growth. Online shoppers are frequently perplexed by e-commerce websites, their products, quality, and services. Customers make purchasing judgments based on social media reviews. The most popular product will, of course, be the most in demand. Merchants and manufacturers can make inventory and service level decisions based on client demand by knowing what products and services they want. Customers can make purchase decisions at both the product and E-Commerce website levels. Opinion mining on realtime data is a difficult task. Every second, an enormous amount of data, such as customer thoughts and feelings, is presented and exchanged on Twitter. These tweets can come in a variety of formats, including passages, pdfs, Figs, audio, and movies. Because of its volume, velocity, and variety, this type of data is referred to as big data. As a result, obtaining real-time data and analyzing it for the purpose of opinion mining is a complex task. Big data is neither un-formatted nor formatted because it is semi-organized and contains a high quantity of data. The Hadoop support also used as a platform in this proposed study to reduce the complexity of big data research.

Customers that shop online have varying opinions regarding the things they purchase. The customer expresses their thoughts, in terms of Reviews, Tweets, likes, ratings, comments, Emoji symbols, and shares. These are all expressions and thoughts data is quite valuable for analyzing opinion in product reviews. However, real-time data gathering, dealing out, and scrutiny are not <u>31<sup>st</sup> October 2022. Vol.100. No 20</u> © 2022 Little Lion Scientific

| ISSN: 1992-8645 <u>www</u>                      | .jatit.org E-ISSN: 1817-3195                        |
|---|---|
| easy. This mission is primarily concerned with  | been used to find the most suited features. To      |
| analyzing Tweets of Amazon products in order to | classify sentences into positive, negative, and     |
| generate product remarks for customers and      | neutral categories, multinomial Nave Bayes has      |
| retailers.                                      | been utilized. Also in [13] Bayesian based learning |

The proposed work is based on Amazon product Twitter data. To provide the greatest and most influence on sentiment analysis, the study incorporates all forms of words and expressions. This suggested project will assist buyers in learning about the popularity and quality of the brand and product. Also aids businesses or organizations in making product, brand, and quality decisions according on customer favorite, feelings, and opinions. The following is how the paper is organized. The related work on opinion mining is presented in Section 2. Section 3 proposes the opinion mining method and materials that will be employed in this proposed rule based system. The planned work's result and comments are presented in Section 4. Finally, in section 5, the conclusion is offered.

# 2. RELATED WORKS

Prediction of consumer product demands has been proposed by Alain Yee Loong Chong et al. [1]. Customers and promotional marketing statistics are available online. Electronic product reviews are analyzed to forecast sales, demand, and service. From the Amazon E-Commerce website, promotional marketing data and customer reviews of electronics products are scraped. Electronics products are subjected to sentiment analysis. Sentiment analysis and influencing factors found from promotional marketing data and online review data are utilized to estimate the number of helpful and negative scores. Machine learning predicts sales or product demand based on influencing factors.

Sentiment analysis using hybrid approach has been advised by Avinash Kumar et al. [2]. In the preprocessing phases, Tweet have been collected and converted upper case to lower case, deleted stop words, URLs, user names that begin with @, and removed ReTweets. TF-IDF and Count Vectorization algorithms were used to extract words. The random forest approach has been utilized. Also in [13] Bayesian based learning system has been developed for detect the subjectivity. The performance has been measured using measures such as precision, accuracy, flscore, and recall. Author has proposed an algorithm that increased accuracy from 58 percent to 75.5 percent. Benito Alvares et al. [3] have been offered sentiment analysis-based opinion mining. The opinion mining has been carried out on movie reviews in this case. The movie reviews are first gathered and saved in a database by crawling ecommerce websites. Then, using POS tagging, the reviews are preprocessed and feature words are extracted. Opinion mining has been devised by Mahalakshmi R et al. [5]. The Hadoop Flume tool has been used to collect Twitter data, which is then stored in HDFS and analyzed. By deleting superfluous words, characteristics are extracted. The MapRedcue function is used to calculate polarity checking, which is used to do sentiment analysis. MongoDB stores each Tweet as well as its polarity state. Finally, using the R tool, the polarity score is shown.

Opinion Mining on using Lexicon-Based Sentiment has been proposed by Dibakar Ray et al. [6]. Negation words, slang terms, multiple punctuations in a word, egree modifier words, repeated letters in a word, and capitalized words were all employed in the analysis [4]. In [12] have been offered the multi sentiment lexicon based system has been approached. It has had a positive effect on Sentiment Analysis. In Indonesian literature, Valmeekam Karthik et al. [7] have been supported using Machine Learning Classifiers, Convolutional Neural Networks, and Artificial Neural Networks to perform sentiment analysis [8] on emojis (facial expression). Also in [14] Genetic Algorithm based Support Vector Machine mechanism [16] has been suggested for optimize performance of Indonesian opinion the classification. Emoji-free tweets are deleted. Unicode is used to convert all emojis. When no emoji are present, all slash symbols and Xs in tweets are erased, and other characters are

#### Journal of Theoretical and Applied Information Technology

31<sup>st</sup> October 2022. Vol.100. No 20 © 2022 Little Lion Scientific



ISSN: 1992-8645 www.jatit.org converted to ASCII code before being translated to Unicode. In [15] author advised multi-labeled mechanism for emoji classification using mutual data based feature selection approach. Rini Wongso et al. [9] have been offered opinion investigation using TF-IDF. Web crawling is used to collect 5000 news articles, each of which is classified by politics, technology, economy, sports, and health. Text has been preprocessed by ignoring stop words, meaningless terms, and lemmatizing. TF-IDF (Term Frequency-Inverse Document Frequency) selects feature words to estimate the SVD (Singular Value Decomposition) for the Nave Bayes method to segregate the News articles. In [17] Text has been analyzed by the sentiment analysis task furthermore state of the art also provided for the literature reviews.

The mixture of ReTweet and Likes in opinion mining has been proposed by Rizal Setya Perdana et al. [10]. It analyzed tweets using the Nave Bayes classifier, the ReTweet count, and the likes count [11]. The Tweets are categorized as positive or negative in order to calculate the f1 score, which is then used for further analysis. Customer's opinions on E-Commerce products have been established as a result of these literature reviews.

This suggested study identified that existing research work on opinion mining have some limitations such as limitated features have been used, besides limitated polarity words have been used. Generally if the opinion word is not present in the opinion word corpus or dictionary then the word simply ignored. This can impact on the overall classification and customer's opinion can be reversed. This problem is ignored by this proposed work, here huge polarity words opinionated words utilized. And for classification and opinion prediction purpose a new rule-based opinion mining approach developed. This approach can reveal specifics of well-known products and their quality, and sellers can learn about their customer's opinions and expectations about their product.

#### atit.org E-ISSN: 1817-3195 3. METHODOLOGY OF THE PROPOSED WORK

In this proposed work four different kinds of dataset or Tweets about Amazon products have been incorporated namely Electronics Items, Fashion Items, Mobile Products, and Books. The Twitter API has been utilized in this suggested work to collect real-time data from Twitter concerning E-Commerce items. Tweets about Amazon products have been gathered in real time via tweets for this suggested mechanics. Natural Language Processing (NLP), Twitter API, Python Libraries, and Hadoop have been incorporated to examine and visualize the result. Rule Based Classification Technique has been developed and employed in this suggested study for opinion mining and opinion classification. Opinion mining is the process of discovery and mining the opinion words from a created database. Opinionated terms were found, extracted, and analyzed in this proposed system. To extract and analyze opinion terms, the following general procedures have been employed.

- Tweets about Amazon products have been collected with facilitate of Twitter API and Python.
- Collected Tweets have been stored in HDFS
- Collected Tweets have been preprocessed by the methods of unwanted symbols removal, tokenization, stop words removal, stemming, and POS-Tagging.
- In order to create multi polarity words this proposed work developed the sentiment score calculation method has been created by the integration of Vader sentiment, Text Blob, AFINN, Newly created Corpus, and SentiWordNet Methods.
- Preprocessed Tweets have been incorporated in opinion mining tasks namely polarity computation for each opinion word and sentiment score computation for each Tweet.
- Rules have been generated for opinion classification and prediction.

<u>31<sup>st</sup> October 2022. Vol.100. No 20</u> © 2022 Little Lion Scientific

| ISSN: 1 | 992-8645 <u>www</u>                       | E-ISSN: 1817-3195                                  |
|---------|---|--|
| *       | Based on the generated rules opinionated  | been classified as positive Tweets and 30 Tweets   |
|         | Tweets have been classified as positive   | have been classified as Negative Tweets. Out of    |
|         | Tweets and negative Tweets.               | preprocessed 99833 numbers of Tweets on            |
| *       | Then based on the computed Threshold      | Electronics Items, 40 Tweets have been classified  |
|         | value Tweets have been predicted as       | as Positive Tweets and 98993 Tweets have been      |
|         | positive Tweets and negative Tweets.      | classified as Negative Tweets. Similarly, out of   |
| *       | Resulted classified positive and negative | preprocessed 53667 numbers of Tweets on Fashion    |
|         | Tweets have been stored in MongoDB.       | Items, 53619 Tweets have been classified as        |
| *       | Finally performance of the proposed       | Positive Tweets and 48 Tweets have been classified |

system have been evaluated by the performance analysis approach namely confusion matrix, precision, accuracy, flscore and recall.

#### 4. EXPERIMENTAL RESULTS AND DISCUSSION

For the experimentation purpose 99833 numbers of Tweets on Electronic Items, 53667 numbers of Tweets on Fashion Items, 2684 numbers of Tweets on Mobile Products and 6414 numbers of Tweets on Books have been utilized in this proposed mechanism. After preprocessed the collected Tweets, opinion mining procedure have been applied on the preprocessed Tweets. Then opinion classification has been performed based on the generated rules and opinion prediction has been performed based on the computed Threshold value.

| S.No | Types      | Total<br>Volume | Helpful<br>Tweets | Not<br>Helpful<br>Tweets |
|------|------------|-----------------|-------------------|--------------------------|
| 1.   | Books      | 6414            | 6384              | 30                       |
| 2.   | Electronic | 99833           | 40                | 98993                    |
|      | Items      |                 |                   |                          |
| 3.   | Fashion    | 53667           | 53619             | 48                       |
|      | Items      |                 |                   |                          |
| 4.   | Mobiles    | 2684            | 2678              | 6                        |

Table 1: Classified Opinionated Tweets Volume.

In Table.1, categories represent the involved four different kinds of Tweets about Amazon products namely Books, Electronic Items, Fashion Items and Mobile Products. Total Tweets represents the Tweets after preprocessed, then Positive and Negative Tweets represents the classified positive and negative Tweets from the preprocessed Tweets. From the preprocessed 6414 numbers of Tweets on Books 6384 Tweets have as Negative Tweets. From the preprocessed 2684 numbers of Tweets on Mobile Products 2678 Tweets have been classified as Positive Tweets and 6 Tweets have been classified as Negative Tweets.

Fig.2. illustrates the data mentioned in Table.1. That is Tweets about Amazon product categories namely Books, Electronics Items, Fashion Items and Mobile Products have been preprocessed and classified as positive and negative. These preprocessed and classified volumes of the above mentioned categories have been represented in the graphical form.

From this result Books, Fashion Items, and Mobile Products have been achieved higher positive volume in classification. But Electronic Items only achieved lower positive volume higher negative volume in classification.

Table.2. describes the proposed rule based classification and prediction mechanism achieved performance analysis report on Amazon product categories namely Books, Electronic Items, Fashion Items, and Mobile Products. For all the four different dataset the proposed mechanism has been achieved above 90% accuracy, precision, recall and fl-score. This values sounds that proposed mechanism has been performed well with reasonable values. Fig.3. illustrates the proposed opinion mining approach attained performance analysis report on four diverse kinds of Amazon categories represented in Table.2.

# 5. CONCLUSION

The objective of this study is to help the business people to find the customer's opinion effectively by improves polarity words or opinion words finding methods. This objective has been

#### Journal of Theoretical and Applied Information Technology

<u>31<sup>st</sup> October 2022. Vol.100. No 20</u> © 2022 Little Lion Scientific TITAL

ISSN: 1992-8645 www.jatit.org achieved by the proposed methodology. This has been proved by the performance analysis report. Buyer's opinions of Electronic Items, Fashion Items, Mobile Products, and Books on Amazon E-Commerce have been collected in the form of Tweets from Twitter in this suggested work. The proposed rule-based method has been used to assess variations in public opinion about various items on Twitter, categorizing them as positive or negative, and predicting buyer's view. Polarity scores have been generated for each opinionated phrase, and sentiment scores have been obtained for each Tweet, using the described sentiment score computation approach. Opinion has been separated and projected based on the expected Threshold

atit.org E-ISSN: 1817-3195 value. This work suggests that the proposed work performed well and created greater accuracy based on the results of this proposed effort. As a consequence of this research, it appears that Amazon's Fashion Items, Mobile Products, and Books have a higher positive than negative opinion rate. These three products appear to have been well received and suggested by the buyers. However, the favorable opinion rate for electronic items is lower than the negative opinion rate. This indicates that Amazon Electronic Items are not well-liked or recommended by purchasers.



Figure 1: Architecture of the Proposed Work

# Journal of Theoretical and Applied Information Technology <u>31<sup>st</sup> October 2022. Vol.100. No 20</u> © 2022 Little Lion Scientific



Figure 2: Classified Tweets using Proposed Technique

| S.No | Categories       | Accuracy | Precision | Recall | F1- Score |
|------|------------------|----------|-----------|--------|-----------|
| 1.   | Books            | 92.94 %  | 100%      | 92.90% | 96.32%    |
| 2.   | Electronic Items | 97.84%   | 100%      | 97.77% | 98.21%    |
| 3.   | Fashion Items    | 94.37%   | 100%      | 94.36% | 97.10%    |
| 4.   | Mobiles          | 98.40%   | 100%      | 98.39% | 99.19%    |

| Table 2. Doufour an on | an almaia Domout  | on Duonogod | Classification | Tachaiana |
|------------------------|-------------------|-------------|----------------|-----------|
| Table 2: Performance   | anaivsis kepori   | on Proposea | Classification | recontaue |
|                        | many and any part |             |                |           |



Performance Matrics

Figure .: Performance Analysis Report for Proposed Work

### LIMITATION AND FEATURE DIRECTION

In this proposed work involved the simple rule based classification model for classify the positive and negative tweets. In addition imbalanced data has been involved in this work. In feature study balanced data will be take in account and artificial neural network or deep learning model will be utilized to classify and predict the product opinion about the customer's.

# **REFERENCES:**

- Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*, 55(17), 2017, 5142-5156.
- [2] Kumar, A., Sharma, S., & Singh, D. Sentiment Analysis on Twitter Data using a Hybrid Approachl. *International Journal of Computer Sciences and Engineering*, 7.
- [3] Alvares, B., Thakur, N., Patil, S., Fernandes, D., & Jain, K. Sentiment analysis using

opinion mining. International Journal of Engineering Research, 5(04), 2016.

- [4] Maheswari, S. U., & Dhenakaran, S. S. Sentiment analysis on social media big data with multiple tweet words. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN*, 2019, 2278-3075.
- [5] Mahalakshmi, R., & Suseela, S. Big-SoSA: social sentiment analysis and data visualization on big data. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(4), 2015, 304-306.
- [6] Ray, D. Lexicon Based Sentiment Analysis of Twitter Data. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 5, 2017, 910-915.
- [7] Karthik, V., Nair, D., & Anuradha, J. Opinion mining on emojis using deep learning techniques. *Procedia computer science*, 132, 2018, 167-173.
- [8] Khan, F. H., Bashir, S., & Qamar, U. TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision* support systems, 57, 2014, 245-257.

<u>31<sup>st</sup> October 2022. Vol.100. No 20</u> © 2022 Little Lion Scientific www.jatit.org



[9] Wongso, R., Luwinda, F. A., Trisnajaya, B. C., & Rusli, O. News article text classification in Indonesian language. *Procedia Computer Science*, 116, 2017, 137-143.

ISSN: 1992-8645

- [10] Perdana, R. S., & Pinandito, A. Combining likes-retweet analysis and naive bayes classifier within twitter for sentiment analysis. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10(1-8), 2018, 41-46.
- [11] Uma Maheswari, S., & Dhenakaran, S. S. Opinion exploration of tweets and amazon reviews. *Int. J. Sci. Res.(IJSTR)*, 2020, 1-9.
- [12] Yang, X., Zhang, Z., Zhang, Z., Mo, Y., Li, L., Yu, L., & Zhu, P. Automatic construction and global optimization of a multisentiment lexicon. *Computational intelligence and neuroscience*, 2016.
- [13] Chaturvedi, I., Ragusa, E., Gastaldo, P., Zunino, R., & Cambria, E. Bayesian network based extreme learning machine for subjectivity detection. *Journal of The Franklin Institute*, 355(4), 2018, 1780-1797.
- [14] Prastyo, P. H., Ardiyanto, I., & Hidayat, R. A Combination of Query Expansion Ranking and GA-SVM for Improving Indonesian Sentiment Classification Performance. *Procedia Computer Science, 189*, 2021, 108-115.
- [15] Ahanin, Z., & Ismail, M. A. A multi-label emoji classification method using balanced pointwise mutual information-based feature selection. *Computer Speech & Language*, 73, 2022, 101330.
- [16] Khanday, A. M. U. D., Khan, Q. R., & Rabani, S. T. SVMBPI: support vector machine-based propaganda identification. In *Cognitive Informatics and Soft Computing*, 2021, (pp. 445-455), Springer, Singapore.
- [17] Kalaivani, M. S., & Jayalakshmi, S. Text-Based Sentiment Analysis with Classification Techniques—A State-of-Art Study. In Computer Networks and Inventive Communication Technologies, 2022, (pp. 277-285) Springer, Singapore.